Advances in Educational Technology and Psychology (2020) 4: 68-75
Clausius Scientific Press, Canada
DOI: 10.23977/aetp.2020.41011
ISSN 2371-9400

An Empirical Analysis of Data-driven Intelligent Teaching Based on Cloud Class Platform

Xinhong Liu[1,*], Yuan Feng[1], Chunxia Wu[1], Yaqin Lu[2] and Di Gao[3]

[1]*Department of Mathematics and Physics, Beijing Institute of Petro-chemical Technology, Beijing102617, China*

[2]*Department of Foreign Language, Beijing Institute of Petro-chemical Technology, Beijing102617, China*

[3]*School of Mechanical Engineering, Beijing Institute of Petro-chemical Technology, Beijing102617, China*

*Corresponding Author

Keywords: data driven, education evaluation, cluster analysis, correspondence analysis, harmonic curve

Abstract: At present, with the vigorous development of education platform, it is necessary to evaluate the effect of teaching process based on the data on the platform, but the evaluation method is relatively simple, lacking the data-driven evaluation based on the learning process. In order to achieve data-driven learning evaluation and improve teaching efficiency, the cloud class platform is used to collect the data of learners' learning process and the analysis and research of data visualization are carried out based on the platform. First, the students are divided into four categories by harmonic curve and cluster analysis. Second, the advantages and disadvantages of all kinds of students are pointed out by correspondence analysis. These results can guide teachers to use education data to mine the relevance between achievements and knowledge points, and help teachers to implement the talent training strategy of individualized teaching, improve teaching quality, provide timely feedback and control for students' learning, facilitate students' independent learning and improve learning effect.

1. Introduction

With the advent of the era of "Internet + education", great changes have taken place in the concept, culture and ecology of education. The education big data has been generated by teaching education in intelligent environment. Using data to drive the reform and innovation of education system has become an important path to improve the level of education development [1-4]. As education researchers gradually realize the importance of education data, making data "voice" becomes the key to play the value of data. The analysis of education data provides effective theoretical and technical support for the study of potential available information of campus data. The results of data analysis can not only reflect the learning effect and future development potential of learners, but also promote people to find out the education state, explore the laws of education

and teaching activities, and guide education and teaching behavior to move towards a more scientific and rational data intensive scientific paradigm [5-7].

With the continuous infiltration of the student-centered education concept [8, 9] and the continuous improvement of higher education level, individualized talent training based on students' aptitude, such as classified management of students, hierarchical training, precise teaching [10] and personalized characteristic development, has drawn educators' attention. However, the existing evaluation methods and quantitative effect of the students' learning process and personalized training are not satisfactory, and there is a lack of evaluation based on students' learning process and personalized data. Data evaluation of students' learning process is the basis of students' learning effect evaluation and individualized education evaluation. Its purpose is to make students understand their learning purpose, advantages and existing problems, so as to stimulate students' potential, promote their development and improve their academic level. At the same time, it also enables teachers to further understand the structure of students' knowledge and ability, so that they can seriously adjust teaching plans, research teaching strategies, and implement individualized education of teaching students according to their aptitude.

At present, in the domestic first-line teaching, the comprehensive evaluation of college students basically adopts the traditional evaluation model, such as the general evaluation score, the average credit score model and so on. Guo Liqin, Chen Li and Deng Kun analyzed the relationship between College Students' performance and ability through the construction of curriculum group model, so as to provide basis for colleges and universities to realize the evaluation of students from comprehensive performance to ability and quality [11]. Chen Xihua, Huang Haining and Huang Peijie use K-Means clustering analysis algorithm, through SPSS modeler software, to find out the distribution of students' scores in each course, and predict the importance of the course, so as to provide basis for the implementation of teaching reform, improvement of teaching quality and learning effect of students [12]. Li Shanshan and Li Quan use factor analysis and cluster analysis to analyze the achievements of students in various courses, evaluate the advantages and disadvantages of students in various disciplines, find out the differences of students' abilities in various aspects, and provide reference basis for teaching reform and student employment guidance [13]. There are human subjective factors in these models, lacking personalized and visual guidance for students, so it is very important to establish evaluation models with fairness, objectivity, scientificity, personalization and visualization.

In view of this, an in-depth analysis of the data collected by the cloud class platform is studied, and the collection, analysis and data visualization are connected to realize data-driven learning evaluation. By using the method of harmonic curve and cluster analysis, the mathematics scores of some major students in our university are classified. Through correspondence analysis, the mathematical scores of college students are classified and corresponding, and the correlation between students' scores and knowledge points is mined out by using educational data. Based on the mining and multivariate statistical analysis of teaching data, teachers can accurately teach and students can accurately learn. It is helpful for teachers to improve the quality of teaching and their academic performance by using data feedback in after class teaching. Through the data of cloud class platform, teachers can understand and determine the knowledge mastery degree and learning needs of different students, so as to give students personalized guidance. The teaching process reflects the transformation of the educational concept from the old three centers to the new three centers, and makes targeted teaching reform to improve the quality of talent training in colleges and universities.

2. Methodology

The research methods mainly include: cluster analysis, harmonic curve and correspondence analysis.

2.1 Cluster analysis

The distance representing the similarity degree between samples is defined. According to the distance, that is, the similarity degree, the samples are classified one by one, and the closely related classes are clustered into a small classification unit, and then gradually expanded, so that the distant relationship is aggregated into a large classification unit, until all the samples are gathered, forming a clustering graph representing the affinity relationship, the basic idea of cluster analysis is to classify variables.

According to the definition of distance, clustering analysis can use the shortest distance method, the longest distance method, the center of gravity method, the class average method or the sum of squares of deviations. In this paper, the sum of squares of deviations is used. This method was proposed by Ward in 1936, and then developed by Orioci et al in 1967. If the classification is correct, the sum of squares of the same kind of samples should be small, and the sum of squares of deviations between classes should be larger. First, each sample is regarded as one class, and then one class is reduced each time. For each reduction, the sum of squares of deviations increases. The choice is to combine the two classes with the smallest increase in the sum of squares until all the samples are grouped into one class. The method of the sum of the squares of the deviation has a good classification effect and is widely used.

2.2 Harmonic curves

The idea of harmonic graph is very similar to Fourier transform, which maps the points in the p-dimensional space to the curve on the two-dimensional plane according to the trigonometric transformation method.

Suppose X_r is the r-th observation for p-dimensional data, $X_r^T = (x_{r1}, x_{r2}, \cdots, x_{rp})$, then the corresponding harmonic curve is

$$f_r(t) = \frac{x_{r1}}{\sqrt{2}} + x_{r2}\sin(t) + x_{r3}\cos(t) + x_{r4}\sin(2t) + x_{r5}\cos(2t) + \cdots, \quad -\pi \le t \le \pi$$

A sample corresponds to a curve.

When clustering, the same curve is close together, and the curves of different classes are separated. It is a very intuitive classification tool. There are two mathematical properties of harmonic curve: (1) keep linear; (2) the square of distance defined by harmonic curve has linear relationship with the square of Euclidean distance, which is the basis of harmonic curve classification.

2.3 Correspondence analysis

Correspondence analysis, also known as association analysis, R-Q-type factor analysis, is a statistical technique for multi-dimensional random variables to reflect the dependency relationship. It shows the relationship between row variables and column variables. It is a visual data analysis method, which is intuitive, simple, convenient and widely used [14, 15].

Correspondence analysis is a low-dimensional graphical method that reflects the relationship between samples (rows) and indicators (columns). By using standardized data and factor analysis

covariance matrix, two most important common factors are extracted. Two common factors are drawn in the same rectangular coordinate system. The relationship between samples and indicators can be investigated according to the distance of data points, mainly the distance between abscissa and scatter point, and the distance of ordinate is not significant for analysis. Package CA [16] in R software is a special correspondence analysis package.

3. Empirical analysis

3.1 Data collection, processing and basic analysis

Online learning platform tracks data and provides data support for learners' learning performance and learning behavior analysis. Based on the online platform of cloud class, the results of 7 tests of Advanced Mathematics A(I) and college entrance examination mathematics (fundamentals) of 56 students are collected as the research objects in this paper. Through preliminary analysis, the difficulty of Chapter 4 and Chapter 5 is increased, the median is low, and the difference is small, as shown in the star chart and box line chart in Figure 1.

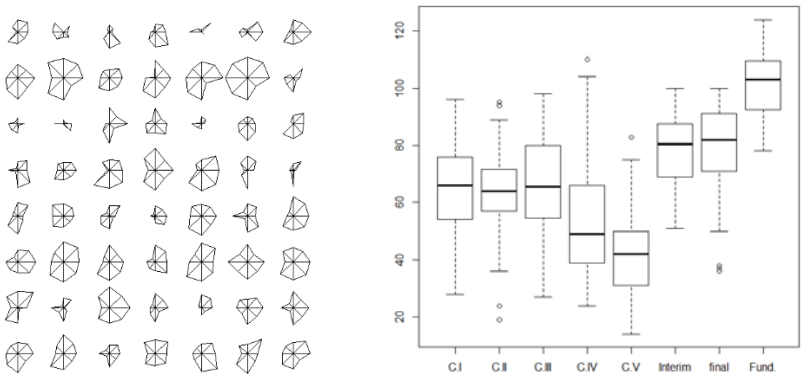

Figure. 1 Star chart and box line chart of basic mathematics and seven test scores

3.2 Establishment and analysis of the model

Similar curves in harmonic curves are twisted together, as shown in Figure 2. 56 students were divided into four categories by the harmonic curve. Students No. 13, 9 and 12 (in the order of increasing student number) were the first category, students No. 43, 55, 30, ·····, 45 were the second category, students No. 2, 5, 16, 15 and 19 were the third category, and other students were the fourth category.

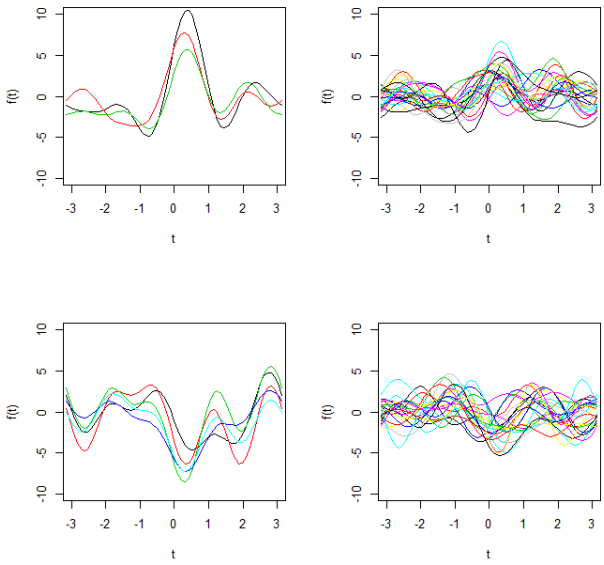

Figure. 2 Harmonic curves of four kinds of students

Cluster analysis is carried out by using the sum of squares of deviation, as shown in Figure 3. There are 3 students in the first category, accounting for 5.4%, which is the smallest proportion. Their scores are relatively balanced, and there is no partial "chapter" phenomenon. They are top students in the class. There are 25 students in the second category, accounting for 44.6%, which is a high proportion. They have good results, play a stable and balanced performance, and occasionally have partial "chapter" phenomenon, but they can basically pass each time. There are 5 students in the third category, accounting for 8.9%, whose scores are relatively poor and they are on the verge of failing. Teachers should urge such students to strengthen self-discipline and improve their learning initiative. There are 23 students in the fourth category, accounting for 41.1%. Different from the students in the second category, the scores of these students are generally not satisfactory and their scores are low. The students have obviously partial "chapter" phenomenon, some chapters have excellent test results, and some chapters fail to pass. If they encounter difficulties, teachers should adjust the pace appropriately, or give guidance to students who are partial to "chapter" after class.

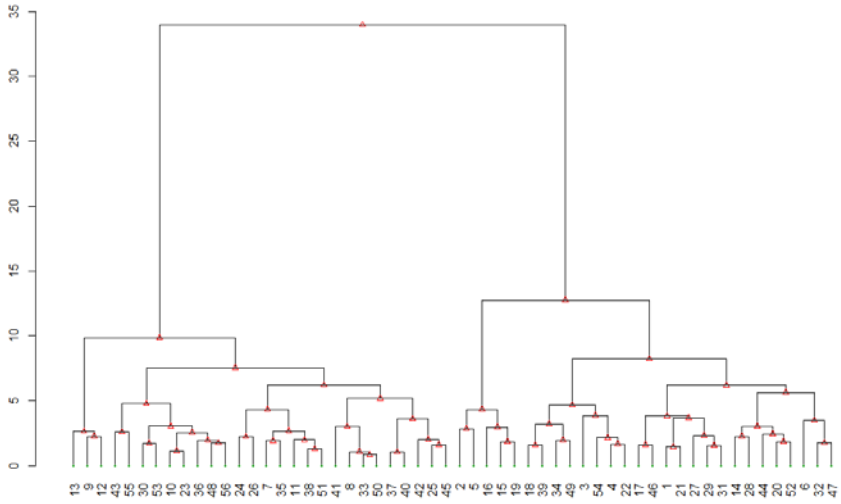

Figure. 3 Cluster analysis

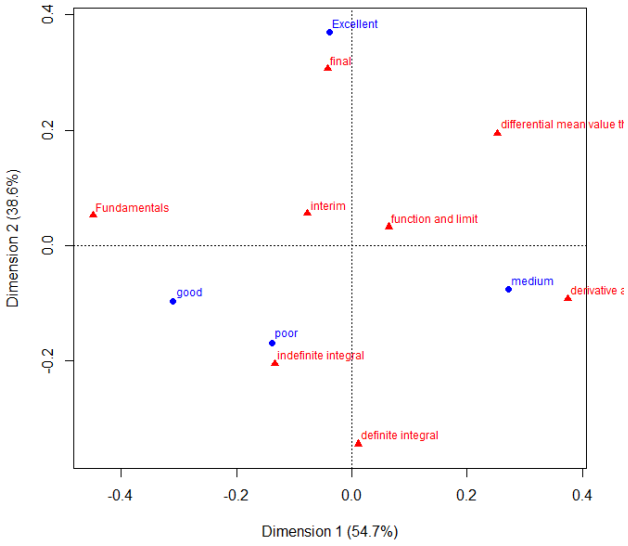

Figure. 4 Corresponding analysis diagram

According to the harmonic curve and cluster analysis, these learners can be divided into four categories: excellent, good, medium and bad. The corresponding analysis scatter chart in Figure 4 is composed of categories and factor coordinates of college entrance examination mathematics (fundamentals), chapters, midterm and final examination. It can be seen from this that the excellent students are not the ones with the best foundation. They perform well in the mid-term, final term, limit and definite integral tests, and their disadvantages are in derivative and differential in Chapter 2. Good students have the best mathematical foundation, but the disadvantage is derivative and differential and differential mean value theorem and its application. After half a year's study, medium students have the worst mathematical foundation and they are good at differential mean value theorem and its application. The disadvantage is indefinite integral, mid-term and final test. The advantage of the poor students is indefinite integral and mid-term and final test. The disadvantage is also the second chapter derivative and differential and the third chapter differential mean value theorem and its application, which completes the initial positioning of the advantages and disadvantages of the students. According to the section of students' disadvantage, teachers should give key guidance, guide students to review key points, and teach students according to their aptitude. It basically forms the general teaching process of precision teaching which consists of data collection, data analysis, teaching analysis, teaching implementation and teaching intervention. In order to help learners achieve maximum learning efficiency, establish evaluation points for each chapter's knowledge points, and achieve targeted teaching, the degree of students' mastering knowledge points from a multi-dimensional perspective should be analyzed, and the unskilled knowledge points based on this should be locked to help students strengthen knowledge acquisition through timely feedback.

4. Conclusion

Through the cluster analysis of 56 students' mathematics scores in college entrance examination (fundamentals), test scores in each chapter, mid-term and final test scores, the students can be divided into four categories. Through cluster analysis and correspondence analysis, students can clearly understand their learning effect and weak links of knowledge mastery, and make up for their own shortcomings in time. According to the results of cluster analysis and correspondence analysis,

teachers can analyze the performance characteristics of four types of students and the mastery effect of learners on a certain chapter, so as to adjust teaching strategies and methods, improve teaching quality, and realize individualized talent training strategies of individualized teaching.

Teachers use quantitative data to reflect the data-driven results based on education and teaching to improve the level of education. Through the data analysis, teachers can master the learning needs of different learners, have insight into the learning behavior path and development law of different learners, so as to carry out personalized teaching and guidance accordingly.

At present, the total number of students participating in the cloud class platform is not large, which limits the sample size of the study, so the reliability and validity of this study need to be further verified in the larger sample. In addition, this study deals with the data of chapter test in the teaching process, neglecting the data of students' learning behavior and so on, which needs further data collection and empirical research. Based on the “Internet+” era, there are more online and offline mixed teaching behaviors and teaching methods. Due to the different requirements of students and the school's network environment, the problems, such as whether this mode has higher teaching level and student satisfaction in other universities, and how to ensure the actual teaching effect, need further empirical research to prove.

Acknowledgements

This work is supported by the municipal URT project "statistics and analysis of big data of learning situation" (2020J00023), the general teaching reform project of Beijing Institute of Petro-chemical Technology "reform and practice of hybrid teaching mode of probability and statistics course" (YB20200802) and "a study on the flipped classroom based on college English autonomous learning mode"(YB20170903).

References

[1] Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: which data, what purposes, and promoting and hindering factors. Teaching and Teacher Education, vol.26, no.3, p.482–496.

[2] Schildkamp, K., Poortman, C. & Handelzalts, A. (2016). Data team for school improvement. School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice. vol.27, no.2, p.228-254.

[3] Van Geel, Marieke, Trynke Keuning, Adrie J Visscher, and Jean-Paul Fox (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. American Educational Research Journal, vol.53, no.2, p.360-394.

[4] Wang Ping (2019). Research on data-driven teaching mode of teachers, Science Innovation. Vol.7, no.6, p.175-180.

[5] Schildkamp, Kim, Maaike Smit, and Ulf (2019). Blossing. Professional development in the use of data: from data to knowledge in data teams. Scandinavian Journal of Educational Research, vol.63, no.3, p.393-411.

[6] Li Xin, Yang Xianmin (2019). The connotation identification and training path of educational data thinking. Modern Distance Education Research, vol.31, no.6, p.61-67.

[7] Wu Yonghe, Tian Yahui, Guo Shouchao, Zhu Lijuan, Ma Xiaoling (2019). Research on the model of learner behavior analysis based on online 3D education platform. China Educational Technology, no.395, p.61-67.

[8] Zhao juming, Gao Xiaohui (2017). Reflections on the implementation of "student-centered" undergraduate teaching reform. China Higher Education Research, no.8, p.36-40.

[9] Zhao juming (2018). Focusing on design: practice and methods (part 2)-a series of studies of the SC undergraduate education reform in the USA (3). Research in Higher Education of Engineering, vol.170, no.3, p.35-50.

[10] Qin Dan, Zhang Lixin (2019). Problems and optimization: realistic review and reflection on practice of precision classroom teaching. e-Education Research, no.11, p.63-69.

[11] Guo Liqin, Chen Li, Deng Kun, et al. (2019). A college student’s capability grading and individual capability analysis based on clustering. Journal of Jiaxing University, vol.31, no.4, p.141-144.

[12] Chen Xihua, Huang Haining, Huang Peijie (2018). Analysis of students’ scores based on cluster analysis. Journal of Qingyuan Polytechnic, vol.11, no.2, p.64-70.

[13] Li Shanshan, Li Quan (2016). Student performance evaluation based on R software factor analysis and cluster

analysis. Computer Programming Skills & Maintenance, no.1, p.41-43.
[14] Cheng Hao, LV Xiaoling, Zhong Yan, et al. (2018). Research on the distribution of APP usage time based on smart phone big data. Mathematics in Practice and Theory, vol.48, no.19, p.158-164.
[15] Zhang Lei, Fang Xu (2018). Application of correspondence analysis in seasonal distribution of multidrug resistant bacteria. Chinese Journal of Disinfection, vol.35, no.6, p.428-431.
[16] Nenadic,O., Greenacre, M. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. Journal of Statistical Software, vol.20, no.3, p.1-13.